3

Deep Extreme Speech

Intimate Networks for Inflamed Rhetoric on WhatsApp

SAHANA UDUPA

"There are some people whom you cannot convince to go beat up somebody," observed Anusha, an Indian journalist. "For them, they [political party campaigners] use religion to sustain the political interest." In an ethnographic conversation in March 2023, the Bangalore-based journalist was describing to us the vast number of WhatsApp groups that political parties have raised in India, and how such groups have lately transitioned from bundled-up masses of members to "specialized and nuanced" clusters. This fine-tuning has led to the parsing and slicing of voting populations based on age, region, gender, lifestyle, and assumed value systems, and offering content to align with such demarcated vectors within specially carved out or embedded WhatsApp groups.

Anusha went on to provide more examples. "Teenagers are primed to talk about violence," she said. "They are given false histories, for example, that Uri Gowda and Nange Gowda [warriors from the Kannada-speaking region in Southern India] killed Tippu Sultan [Indian Muslim ruler of the kingdom of Mysore in the eighteenth century][1] while those of us who have studied the history know it was the British who killed Tippu. Homemakers are treated with temple stories and questions of dharma [righteousness in the Hindu tradition]; for the Marathi-speaking community, content will be on Shivaji [Marathi warrior king who resisted British occupation], and for Kannada-speaking people,

1. The rewriting of history around Tippu Sultan's death sparked a controversy in Karnataka in March 2023, when the state was preparing for regional elections. "'Pure Fiction': Historians on Uri Gowda and Nanje Gowda Killing Tipu Sultan," News18, March 20, 2023, https://www.news18.com/india/pure-fiction-historians-on-uri-gowda-and-nanje-gowda-killing-tipu-sultan-7336759.html.

stories will be on Kittoor Rani Chennamma [Kannada warrior queen who fought British occupation]."

As research participants described to me in various ways, stories of historical and cultural significance are embellished with track beats to raise the level of attraction, especially among younger users who could be drawn to histories beyond the "boring" static formats of textbooks. The varied strands of targeted content are appended to the perception of correct political choices, i.e., the perception that by consuming and sharing certain types of content and by taking part in this sharing practice, one makes the right political choice. In this chapter, I delve into WhatsApp as the unique social infrastructure that makes this form of content sharing an important feature of extreme speech ecosystems, with particular valence and heightened significance in the Global South contexts.

A growing body of scholarship that has examined the entanglements among platform affordances, political propaganda, hateful speech, and disinformation has highlighted the role of encrypted messaging applications in terms of their disinhibiting effects upon users who share problematic posts as well as enabling impacts of in-group camaraderie and content-rich influence strategies (Bursztyn and Birnbaum 2019; Cheeseman et al. 2020; Evangelista and Bruno 2019; Garimella and Tyson 2018; Johns and Cheong 2021; Nizaruddin 2021; Scherman et al. 2022; Recuero, Soares, and Vinhas 2021; chapters in this volume). Qualifying platform-centered analysis, anthropologists and ethnographers have pinned their focus on what people do with media and how complex mediations of lived worlds cluster around, draw upon, and reshape technological possibilities of WhatsApp such as closed-chat architecture and end-to-end encryption (Cruz and Harindranath 2020; Williams et al. 2022). Ethnographic studies on hateful speech and disinformation in the Global South have especially drawn attention to the political use and electoral mobilizations on the messaging service. In Nigeria, studies have shown how WhatsApp differs from algorithmically shaped echo chambers on Facebook or Twitter, prompting politicians to "create partisan environments and inflammatory messages to bolster their candidacy" (Olaniran, this volume). In a study that explicates this point with ethnographic detail, Stalcup (2016) draws attention to styles of sharing on WhatsApp in Brazil that render the messages memorable, leading

to a form of "aesthetic politics" that involves intentional deployment of platform aesthetics to political ends. Studies on WhatsApp in India have argued that WhatsApp groups help to "achieve homophily," preparing a fecund context for microtargeting "not at the individual level but at the group identity level" (Sinha, this volume).

Drawing upon this insightful scholarship, this chapter suggests that WhatsApp's unique role in disinformation and vitriolic ecosystems in the Global South contexts of divisive politics lies not as much in the architectural features of encryption but around particular clusters of social relations it enters, entrenches, and reshapes. In other words, rather than technical design features seen in isolation, WhatsApp might be better understood in terms of interactional and structural dynamics around social relationality, obligation, and kinship (Fortes 2005; Sahlins 2014; Tenhunen 2018), and how social relations across a range of contexts are reified and reproduced through WhatsApp communication, with significant ramifications for political discourse. WhatsApp as a social relational form represents a unique strand in the complex mix of factors that enable speech practices that stretch the boundaries of legitimate speech along the twin axes of truth/falsity and civility/incivility—practices defined here as "extreme speech" (Udupa and Pohjonen 2019).

The key argument is that messaging apps such as WhatsApp alongside domestic social media platforms in regional languages with group functionalities enable what might be described as "deep extreme speech." Deep extreme speech is characterized by community-based distribution networks and a distinct context mix, which both build on the charisma of local celebrities, social trust, and everyday habits of exchange. Deep extreme speech could be seen as the social corollary for technologized deep fakes deployed in political campaigns. This type of extreme speech belongs less in the problem space of truth or the moral space of hatred and unfolds rather at the confluence of affect and obligation, variously inflected by invested campaigns.

The forthcoming sections will develop the concept of "deep extreme speech" first by briefly outlining digital propaganda activities in the Indian context, followed by empirical sections on the two aspects of distribution and content, and a concluding note on policy directions to address deep extreme speech. The chapter builds on ethnographic fieldwork carried out in Bangalore, Mumbai, Delhi, and adjoining towns

since 2013, especially the latest rounds of fieldwork in March and October 2023 among online users, journalists, political consultants, and activists in Bangalore and online interviews with political consultants and journalists, as well as content analysis of WhatsApp groups affiliated with three major political parties during the 2019 general elections.[2]

WhatsApp in Digital Campaigns in India

Digital social media has emerged as a critical election apparatus in India, following the ruling nationalist party's (Bharatiya Janata Party, BJP) pioneering use of social media channels for election propaganda and ideological (re)production in the last two decades. While the party enjoyed significant electoral success in recent years by being the first to capitalize on online networks, it now contends with the emergence of several other political parties vying for similar resources. In an atmosphere of a "permanent campaign" enabled by digital media (Neyazi 2020), political parties across the spectrum have ramped up their digital campaigns across diverse social media platforms. The Indian National Congress (INC), the chief pan-Indian opposition party, has research teams to prepare "counters" to the BJP's public statements and a social media dissemination team that distributes "research data" and composes content for voting publics. Regional political parties such as Samajvadi Party and Shiv Sena are increasingly enlisting the services of commercial political consultants and digital influencers to promote campaign content and leaders. Aside from social media channels attached to political party systems, individual political leaders of all major parties, including regional parties contesting the elections, are now actively recruiting social media teams for campaign work. While these activities indicate more clamor and flux in the online political sphere, such efforts are quite often overwhelmed by the BJP's heavily funded and rapidly adapting campaign structure and its large "volunteer" base. The vastly intricate digital influence operations of the party keep the momentum of party

2. My sincere thanks to Miriam Homer, Neelabh Gupta, Deeksha Rao, Amshuman Dasarathy, and Sudha Nair for their excellent research assistance. This research is supported by the European Research Council funding under the Horizon 2020 program (grant agreement number 714285 and 957442) and the Centre for Advanced Studies Research Group Funding (2023–2025) at the Ludwig-Maximilians-Universität München.

discourse and flows of audiences alive on social media, with war-like spikes in provocation but also routine everyday exchanges that repeat and reproduce the "party line." The party's dominance in the social media space is such a widely acknowledged reality that a major opposition political leader, during a closed-door meeting with academics in 2023 that I attended, remarked in frustration that "the narrative"—public apprehension of the political situation and collective imagination of the shared futures—is now completely, if not irrevocably, in the hands of the ruling party. While social media influence translated into electoral success for the ruling party in the previous elections of 2014 and 2019, the 2024 elections indicated that major factors weighing in on campaign efforts, including the economy, election financing, and welfare promises, rendered social media messaging an intense battleground among different parties competing to settle voter uncertainties in their favor.

The ruling party's digital presence, a dominant campaign structure, has multiple layers and dimensions spanning dissemination and content creation techniques across platforms in ways to dynamically adapt to polymedia environments and how users engage different platforms at different intervals and for different purposes, all while remixing them with newer articulations (Madianou and Miller 2013). Thus, the party has wide uses and campaign deployments across Facebook, X (formerly Twitter), Instagram, Pinterest, and Clubhouse as well as homegrown platforms such as ShareChat, Moj, and Kutumb. For instance, on platforms such as X (formerly Twitter), the party cultivates and draws from argumentative styles of online engagement that typify such platforms, paving the way for "volunteering" warriors and prepped-up tweeters who reproduce the key ideological tenets of the party by bickering with opponents and repeating simplified summaries of what they understand as the nationalist ideology (Udupa 2018a). The BJP's central Information Technology cell—the nodal office for digital influence operations—coordinates a range of top-down campaign activities attuned to different platform cultures. The official line of work is complemented by hired commercial consultants with promises of data-backed influence strategies and a breeding shadow industry that operates through gray practices of clickbait operators, hired influencers, and loosely knit networks of dispersed amplifiers drawn into precarious and informal labor arrangements crafted by ambitious mediators.

With more than 490 million users in 2021 (Degenhard 2023) and as the second-largest online platform for accessing news in India (Newman et al. 2023), WhatsApp has emerged as a major platform for party campaign activities. It figures in the digital campaign structure of the party—and increasingly other political parties—in two prominent ways: creating intrusive channels of distribution and tactical mix of content. Distribution and content aspects of WhatsApp not only feed on the messaging application's unique affordances to create and sustain groups on a subterranean level but also, more importantly, by replicating some of the core aspects of social relationalities.

Intrusive and Intimate Channels of Distribution

The ruling party's use of WhatsApp, developed over the last decade into a stable campaign structure, is a telling testimony for novel forms of distributing ideological and party-favoring content. The party's WhatsApp system consists of multistep distribution with horizontal networks connected to different vertically integrated nodes. The official network channel, headquartered in the capital city of New Delhi, is centrally controlled for disseminating national-level issues; the central IT team dictates the tenor and content of messages in the national space while regional and local units are allowed to compose messages around context-specific issues relevant to respective electoral constituencies.

However, this centricism is strategically positioned in relation to a more dispersed, flexible structure that can draw and retain volunteers. Together, these networks connect "official" workers with other official workers and connect official workers with "general sympathizers" or "well-wishers," allowing sufficient space for official workers to draw "general sympathizers" to become more committed, and general sympathizers to draw other sympathizers and fence sitters. Content flows from node to node. The flexible parts of such networks are not edges of a single core but constitute connected nodes of content building and influence enhancement. Giving an overview of the IT operations of the party, Amit Malviya, the national head of the Information Technology cell of the party, told me in New Delhi that the party has diversified the online communication channels, "in the sense that we have gone down to [regional] states and we are telling them, look, we have to have multiple layers of communication,

we have to communicate at the central level, we have to communicate at the state level, and perhaps localize it even further." The party's strategies around WhatsApp fall squarely within this ambition for distributed networks of influence enhancement.

During a meeting in 2020 in Bangalore, a high-ranking party official in the Southern Indian state of Karnataka was in awe of what they had managed to raise. He exclaimed excitedly, "BJP had formed their own twenty-five thousand WhatsApp groups during the Karnataka [regional] elections. And there were fifty thousand WhatsApp groups during recent Lok Sabha elections in West Bengal [another subnational state]. Only West Bengal! Imagine, fifty thousand times two hundred, that is ten to fifteen lakhs [1 to 1.5 million people]. So, their message has been reaching ten to fifteen lakh people directly. And imagine the forwards which these posts received! This is huge!" He added that the INC, the chief opposition party, had also formed around five thousand WhatsApp groups during the Karnataka elections, and hence he sees "a lot of official WhatsApp groups being formed by party supporters of different parties." The INC runs a "Rahul Gandhi WhatsApp group" on the national level to facilitate direct interactions with the opposition leader (LiveMint 2024).

In 2023, more WhatsApp groups were added and fine-tuned to achieve greater correspondences between group membership and assumed characteristics of members. In the words of one of the interlocutors, "WhatsApp worlds are more streamlined," as several types of strategists, including former journalists, are consulted to "devise plans to reach out to people." During ethnographic conversations in October 2023, journalists turned political consultants added that WhatsApp groups draw upon the sphere of influence that local actors wield within particular communities, building on the social ties of "nano-influencers" (Joseff, Goodwin, and Wooley 2020). Within large residential apartment complexes in urban areas, for instance, influential "uncles"—kin-based authority figures among residents—are encouraged to steer WhatsApp groups for party-favoring narratives, including top-down messaging such as the "Letter from the Prime Minster," which was sent out during the 2024 elections for cascading circulations on the messaging platform (LiveMint 2024).

A vital feature of WhatsApp campaign work is thus the way it helps create intrusive channels for inflamed rhetoric of different kinds. The key motivation for political parties in this regard is to combine top-down

"broadcasts" with "organic bottom-up messaging." They have sought to accomplish this by installing "party men" within WhatsApp groups of family members, friends, colleagues, neighbors, and other trusted communities. "WhatsApp penetration"—defined as the extent to which party people "organically" embed themselves within trusted WhatsApp groups—is seen as a benchmark for a political party's community reach. Local musicians, poets, cinema stars, and other "community influencers" have been recruited to develop and expand such "organic" social media networks for party propaganda. Typically, a party moderator would find his way into a WhatsApp group through local connections or by leveraging community work such as local brokerage to help people to access state benefits and so on, and once admitted, he would relay party content in unobtrusive ways. "WhatsApp groups are intentionally generated," observed Tabassum, a journalist based in Bangalore, and party workers and activists added that right-wing groups meticulously go by the electoral list and start WhatsApp groups for every "sheet [in the electoral list with voter names]."

Such distribution patterns on WhatsApp reveal that while online extreme speech circulation is driven in part by technological features of virality and putative encryption, a significant part of this circulation operates by tapping social trust and cultural capital at community levels, often making deep inroads into the "intimate sphere" of families, kin networks, neighbors, caste-based groups, ethnic groups, and other longstanding social allegiances. Such types of vitriol rely on and rework localized community trust as the key lubricant for the networked pipeline of extreme speech and hate-based disinformation.

To be sure, trust and mistrust are socially negotiated, and not all content that comes on WhatsApp groups is likely to be trusted at once. Kishor, an activist in Bangalore, recounted his experience of interacting with school children in Shimoga, a semiurban area in Southern India. "Schoolchildren who were studying in the Kannada [regional language] medium schools said they were aware of algorithm-based tweaking and that social media narratives have a right-wing bias. 'WhatsApp University' as a derogatory term to signal deliberate creation of distorted knowledge is also common among users. Yet, we are not able to break the cycle of networks sharing such content." As philosopher Hartmann (2011) suggests, trust and mistrust are not mutually exclusive categories;

they evolve dynamically. For instance, anthropologist Beek's (2018) ethnography on "romance scammers" in Ghana illustrates that people on online dating sites engage in romantic relationships with their contacts despite mistrusting them. In other words, social practice exceeds subjective assessments of trustworthiness.

WhatsApp content flows are also not without friction. Sample this conversation between a political consultant and Deeksha, our research assistant in Bangalore, during one of the interviews:

> GM: I am speaking as an individual now, not as a political consultant. I used to receive a lot of forwards [on WhatsApp] from my own family members. So, I used to have debates, discussions, everything with them.
> DR: Do you finally do what I did? I left the group for my own sanity.
> GM: So, in the beginning I used to do that. But later, I thought let us give it back to them.
> DR: So you were a fight person. I was a flight person. So, it's good to know someone is fighting battles.
> GM: Yes, yes, we have to fight.

The resolve to fight or the urge for flight notwithstanding, the conversation attests to the deep inroads that partisan content has managed to make in and through messaging applications like WhatsApp. "There are often intense political battles within families," said Tabassum about her own experiences. "My uncle fumes at me for toeing the line of the ruling party, and I hold ground despite pressure."

Simulating the Social

If community trust embedded in WhatsApp eases the flow of extreme speech, the manner of crafting content for electoral and political influence is no less significant. To examine content patterns on partisan WhatsApp groups, we carried out content and thematic analysis of a corpus of messages sourced from eight WhatsApp groups. The groups were affiliated with or explicitly supported three major political groups: nationalist BJP (groups identified here as B4I, JSR, and ABP); the opposition Indian National Congress (KLI, Inspire, Align) and Aam Aadmi Party, an urban political party and the ruling party of Delhi (AAP). Our

research assistants joined the groups either using the phone number publicly announced by the parties with an invitation to join the groups or by obtaining oral consent of the moderators. Data was gathered during the 2019 general elections, which were held between April 1 and May 19, 2019. The length of the data-gathering period varied for different groups and ranged between September 2018 (oldest) to January 1, 2020 (latest) for the sampled election period. The messages, which contain text and media (still images, GIFs) in English, Hindi, and mixed registers of Hindi and English, were qualitatively analyzed by identifying thematic categories and tones through bottom-up coding, and with a binary classification for extreme speech and lists of types and targets of extreme speech (Tables 3.1–3.4). Weblinks in the messages were not included in the coding. A total of 30,887 messages were coded (*KLI* 9,550 text; *ABP* 4,701 text, 495 media; *AAP* 4,216 text, 281 media; *B4I* 3,583 text, 394 media; Inspire 1,893 text; *Align* 924 text, 67 media; *JSR* 658 text, 280 media).

Table 3.1: Overview categories

Category Key	Category Key Name
1	Nationalism/patriotism
2A	Religion (Hindu)
2B	Religion (Islam)
2C	Religion (Christianity)
3AA	Politicians (local, regional) NATIONALIST
3AB	Politicians (local, regional) INC
3AC	Politicians (local, regional) Other
3BA	Politicians (national) NATIONALIST
3BB	Politicians (national) INC
3BC	Politicians (national) Other
3C	Politicians Modi
3D	Politicians Rahul Gandhi
4	Development
5	Personal wellness/greeting
6	Historical
7	Party symbols
8	Any other
9	Inhuman/violence

Table 3.2: Overview of tones	
Tone Key	Tone Key Name
1A	Sarcasm (with humor)
1B	Sarcasm (without humor)
2	Informational
3	Greeting/personal
4	Confrontational
5	Graphic
6	Warning
7	Allegation
8	Soothsaying
9	Eulogizing
10	Any other

Table 3.3: Extreme speech types	
Extreme speech Key	Extreme speech Key Name
1	Offensive to community
2	Call for violence
3	Violence through reference
4	Call for exclusion from nation/community
5	Sexist
6	Casteist

Table 3.4: Extreme speech target groups	
Target of Extreme speech Key	Target of Extreme speech Key Name
1	Muslims
2	Women
3	Dalits
4	"Pseudoliberals" (progressive-liberals)
5	General public/unmarked audience
6	Right-wing groups

A quantitative summary analysis provides an overview of the distribution of thematic categories and extreme speech types across three political groups, which were calculated as a percentage of the total number of messages per group to highlight the significance of different content types within the group. In terms of message types, AAP and nationalist WhatsApp groups relied more on images in comparison to the Congress WhatsApp groups.

Across all the groups, an interesting finding is that a very small number of users sent the greatest number of messages, confirming other studies that have documented the prominence of "super users" in online political networks, who "account for the vast majority of posts and of extremist language" (Kleinberg, van der Vegt, and Gill 2021). Figure 1 shows the twenty most active users based on the number of total messages (including text messages, media, and web links) they shared. The users were ranked from 1 to 20, with 1 referring to the most active user. Many groups had a single user who was significantly more active than their fellow top users, which is especially evident in the case of *ABP* (the number of messages sent by the first and second most active user differs by more than half: sender 1: 1212 messages; sender 2: 507 messages). This pattern is visible in other groups as well, although less starkly in *AAP* (sender 1: 789; sender 2: 570), *KLI* (sender 1: 773; sender 2: 630), *Align*

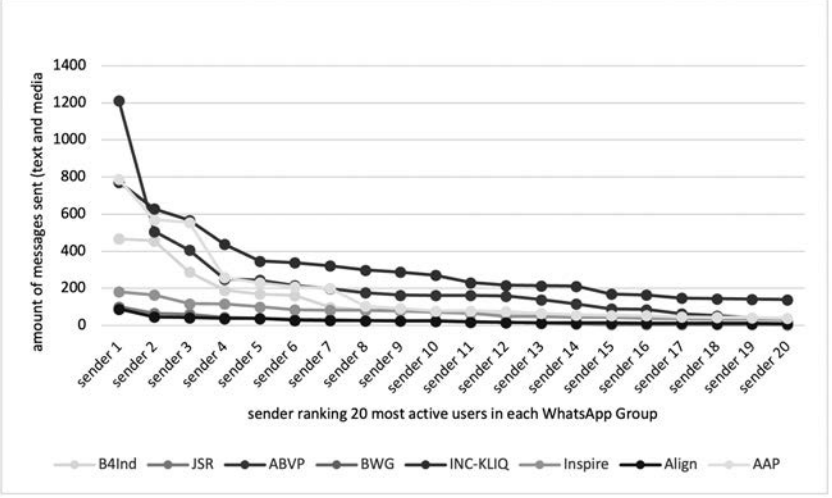

Figure 3.1: The twenty most active users in WhatsApp groups. Source: Author

(sender 1: 89; sender 2: 47), *Inspire* (sender 1: 182; sender 2: 166), and *JSR* (sender 1: 102; sender 2: 66). This illustrates the "long tail" phenomenon in online political groups, especially within extreme speech scenarios, highlighting how discussion communities are sustained largely by the overactivity of a handful of ringleaders with a long tail of relatively underactive followers and signaling the possibility of sponsored or highly motivated actors.

Themes

Text messages and media were analyzed based on eighteen thematic categories, and up to three categories were included in a single message if multiple strong thematic indicators were found. The residual "any other" category included empty messages, advertisements for consumer products and services, notifications of members joining or leaving the group, and others. Thematic categories were mapped across three ideological/political groups (nationalist, Congress, and AAP).

Discussions around politicians were a core thematic strand across all the groups. Prime Minister Narendra Modi was the most widely cited politician (nationalist: 760 times; Congress: 862 times; AAP 415 times). In contrast, mentions of Rahul Gandhi were fewer (nationalist: 242; Congress: 383; AAP 28). Regional politicians also featured across the groups, revealing the locally specific concerns of the groups.

Nationalist WhatsApp groups stood out for their higher prevalence of discussions around religion. More than 10 percent of the messages were about positive portrayals of Hinduism (720), followed by references to Muslims (314 messages). Personal wellness and pleasant greetings were more common in these groups (950 messages) while the Congress groups had 659 messages in this type (a smaller percentage of the total messages within the group compared to nationalist groups). As a percentage of the total number of messages in the group, development topics featured most frequently in *AAP* (153), followed by nationalist (269) and Congress (84) groups. Historical themes were used within nationalist groups (194) and Congress (166) groups, but only sparsely in AAP (38). Party symbols were more common in nationalist groups (197), and so was the common symbolism of "Jai Shree Ram" (chant for the Hindu god). Patriotic themes were prominent in the nationalist group (556),

far more than Congress (180) and AAP groups (123). Violent topics were more common in nationalist groups (329), followed by Congress groups (138) and AAP (75).

Extreme Speech

A binary classification for extreme speech that depicted offense, exclusion, or violence revealed that only less than 3 percent of the total messages belonged to this category. The highest occurrences were in the nationalist groups (2–3 percent), while they were less than 1 percent of the total message instances in the other two groups. These messages were further coded for six types of extreme speech (Tables 3.3 and 3.4). In the smaller volume of extreme speech instances in AAP and Congress groups, the messages largely comprised offenses to communities (Congress: 18; AAP: 5). Align, in comparison to other Congress groups, had one instance of extreme speech that was casteist and four instances of calls for an exclusion from the nation or community. AAP also showed one instance of extreme speech that called for violence.

Among nationalist groups, more than 20 percent of extreme speech instances (53) had calls for violence and about 10 percent (27) used acts of violence as a justification for violence. They also had instances of casteist (1) and sexist (3) speech. Aside from offenses to communities (57), calls for exclusion from the nation or communities (72) were common in these groups.

Six different targets of extreme speech were considered for analysis (Tables 3.4). In the case of AAP, extreme speech was largely targeted at right-wing groups (4). Others were toward a general, unmarked audience (2). Similarly, extreme speech within KLI (8) and Inspire (4) groups were directed toward right-wing groups. Extreme speech in the Align group was toward general, unmarked publics (5), with one case of extreme speech toward "pseudoliberals" (a derogatory term for progressive liberals). The most common target group for extreme speech among nationalist groups was Muslims (65 percent), and the remaining messages were targeted at general, unmarked publics (53), "pseudoliberals" (18), and women (2).

The mix of content observed in the groups prompted an inquiry into the temporal flow of different types of content. When content types

were plotted sequentially, the graph revealed an interesting pattern of shifting themes and tones, which oscillated between provocative and pleasant content (see Figure 2 to see this pattern in an exemplary nationalist WhatsApp group). Combined with ethnographic interviews, this message flow pattern indicates how party workers, once admitted into WhatsApp groups, would relay party messages in unobtrusive ways, often embellished with jokes, good morning greetings, religious hymns, microlocal municipal issues such as water or electricity supply, and other kinds of socially vetted and existentially relevant content. The temporal flow of such messages—amplified and articulated by ordinary users—is characterized by the sudden appearance of explicitly hateful messages against Muslims in the midst of an otherwise benign sequence of pleasant or "caring" messages. The flow of content thus simulates the lived rhythm of the social.

While one interpretation of the "simulation of the social" points to deliberate attempts at camouflaging the context of extreme speech dissemination, another interpretative frame would be to recognize the normalizing effects it can have upon everyday WhatsApp conversations. There is an attempt to "dilute the context of delivering extreme messages," observed an interlocutor. "That is how you normalize." Considering that just about 3 percent of the total messages on WhatsApp groups in this study were found to be extreme, WhatsApp content mix and temporal flow reveal unobtrusive ways of embedding divisive content by re-creating the familiar worlds of everyday social exchange.

Deep Extreme Speech

Intimate networks of distribution and socially vetted content flows discussed in the preceding sections highlight the specificity of WhatsApp groups as a communication form in shaping the ecosystems of extreme speech and partisan political messaging more broadly. The embedding of this type of content especially within nationalist WhatsApp groups shows how social trust and context camouflaging have emerged as key aspects of disseminating and normalizing contentious content, leading to what I describe as "deep extreme speech."

Deep extreme speech centers community allegiances in distribution logics, framing extreme content as part of the everyday. With distinctive

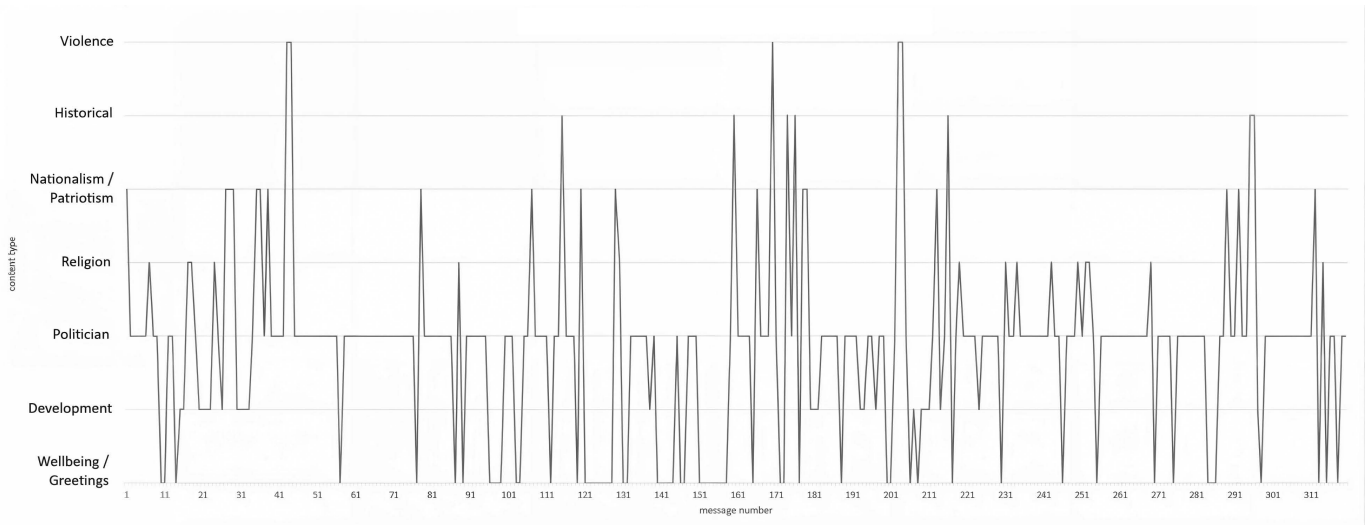

Figure 3.2: Chronology of content in a nationalist WhatsApp group (3,583 messages). Source: Author

distribution patterns, content mix, and temporal flow, such forms have placed extreme speech at the confluence of affect and obligation, thereby delinking it from the impersonal constructions of truthfulness or moral constructions of hatred. In other words, when messages are embedded within personalized, trust-based networks, what ensues is not as much a problem of truth (whether it is true or false) or a problem of morality (whether it is good or bad) but an emotional or obligatory urge to share them and be in (if not with) the flow.

The affective dimension of extreme speech, for instance, is starkly evidenced by the fun cultures of online exchange when people who peddle exclusionary discourse take pleasure and celebrate their collective aggression (Udupa 2019). In the context of deep extreme speech, fun cultures are amplified by social trust and the familiar language of in-group exchange. The aspect of obligation is pronounced in the case of deep extreme speech, as users within kin or kin-like networks feel the need to share and respond to the messages they receive. Any form of inaction on received content conflicts with the sense of obligatory ties and reciprocity that define socially thick networks of deep extreme speech. In such circulatory milieus, responsible action is itself conceived as circulation—the sense that by forwarding the messages one has done one's duty. These social relational dynamics reconfigure trust in extreme speech contexts, as intimate networks become imbued with the intensities of political discussions.

Deep extreme speech that works its way through intimate channels of kin and kin-like relations amid a tactical mix of content might be seen as the social corollary for technologized deep fakes and commercial digital influence services. Analytically distinct but intermingled in practice, these forms are reconfiguring, rather than dismantling, structures of trust in political discourse.

While offering a vital infrastructure for deep extreme speech, WhatsApp communication has ramifications for political discourses beyond the seemingly self-contained worlds of in-group conversation. "There appears to be some sort of coordination across WhatsApp groups of extreme actors since politically relevant contentious content jumps between them in quick time," observed two activists in Bangalore, offering a perspective from their daily navigations of online discussions and interactions with user communities affected by hateful speech.

Moreover, WhatsApp sustains cycles of circulation around specific incidents that would otherwise have disappeared from public memory as a singular incident. "Take the case of a hijab-wearing student who was yelled at by a college lecturer in Karnataka,"[3] said one of my interlocutors. "Someone filmed this chiding, and the video started circulating on WhatsApp. A one-off exposure to politically charged extreme conduct then becomes a prolonged experience for the victim."

Studies have also gathered evidence for the ways in which the ruling political party manipulated social media narratives by using thousands of WhatsApp groups with dispersed volunteers and "loosely affiliated online supporters" to engage in "trending" campaign-friendly hashtags on Twitter (Jakesch et al. 2021). Through such "cross-platform media manipulation" tactics, "hundreds of trends were fabricated" during the 2019 elections (2021). These trends were later picked up by other media outlets, contributing to coordinated amplification of the ruling party's campaign line. By linking content across platforms, therefore, digital influence strategies have sought to evade the limits of encryption and boundaries of closed communities, embedding socially sanctified content derived from WhatsApp communication within multiple streams of digital discourse.

Patterns of deep extreme speech are noticeable in other parts of the Global South, as studies in Bangladesh, Indonesia, Kenya, the Philippines, Brazil, and Turkey, attest (Ong and Cabañes 2018; Saka 2018). In these contexts, large networks of actual, real-world people peddle and amplify animosities online, raising a vast "human infrastructure" for disinformation (Olaniran, this volume). Actors who moderate and lead WhatsApp groups are not highly visible Twitter celebrities but represent smaller nodes that are aplenty and dispersed. During interactions in the AI4Dignity project, a collaborative coding project for AI-assisted content moderation I steered, Gilberto Scofield, a leading fact checker in Brazil, described such community level partisan digital actors as "hyperlocal influencers." They are tasked with "maintaining" about two thousand people on different WhatsApp groups and send out "very useful and local messages," similar to Hindu nationalist moderators

3. Divya Arya, "Karnataka Hijab Controversy Is Polarising Its Classrooms," BBC, February 15, 2022, https://www.bbc.com/news/world-asia-india-60384681.

conveying helpful information on the local water supply and road development. Importantly, such hyperlocal influencers live in the same areas as targeted communities and have a "good understanding of organized crime militia and political parties." These "independent guys" also from the favelas are seen as the "voice of people." The content they provide is seen as "quality information."

Conclusion

WhatsApp's role as a key conduit for deep extreme speech in diverse contexts underscores the need for directing regulatory and policy efforts at the manifold impacts of the vastly popular messaging service. While content moderation is without doubt a significant regulatory measure but hard to implement for encrypted messaging, the analysis presented here stresses the need for a renewed focus on networks of actors and practices as they develop on-ground in closed loops but can scale with active digital influence strategies. Aside from platform governance measures around design features, moderation, transparency, and accountability, which also need to be monitored for potential regulatory excess by state agencies as well as regulations in campaign financing, actor-focused measures would involve regulating the practices of precarious information warriors and data brokers and their role as dispersed amplifiers, curators, and creators of content within WhatsApp networks.

In a recent concept note on online hate speech, the European Commission (2017) observed that while cooperation between governments and civil society organizations is necessary, "cooperation of IT companies with civil society also plays a fundamental role in counternarrative campaigns." According to the note, there is evidence that deploying social media advertising tools to "target audiences" improves awareness and engagement and leads to a substantial increase of NGO's social media presence. To this policy direction, it might be added that civil society and industry should actively involve local WhatsApp groups already constituted by political parties by crafting organic interventions that can *repurpose* existing groups. Such measures will also be crucial to track evolving tactics of digital influencers, such as the instrumentalization of "status update" and other newly introduced features

of WhatsApp, including ways to circumvent the barriers the company has raised in terms of capping the number of forwards and labeling forwarded content by spawning multiple human networks.

Convening self-styled political trolls, local politicians, and commercial digital influencers for awareness-raising activities and sensitizing them about global human rights standards and the dangers of digital campaign manipulations is a necessary step, and so are efforts to strengthen grassroots anti-hate communities to report online extreme speech to social media companies and monitor progress once complaints are raised. Strengthening local communities to petition lawmakers is another important measure, while concurrently supporting local organizations and groups in establishing hate-monitoring dashboards constitutes an equally significant endeavor. In the Global South context, a multimedia strategy is crucial in ways that community trust as the key lubricant for extreme speech is recalibrated through television, radio, and other popular media, and the epistemes of the "WhatsApp University" can remain substantively contested rather than serving as the new normal.